

door observatie in de klinische praktijk

P. Ram, S. Hobma, K. Aretz

Samenvatting

Inleiding: Een goede beoordeling van medische competentie vereist verschillende toetsmethoden. Er moeten keuzes worden gemaakt: wat wordt op welk niveau en op welke wijze getoetst? De te toetsen taken en de opleidingsfase bepalen de keuze van de toetsmethode.

Blauwdruk: Belangrijke aspecten van de keuze van een toetsmethode zijn samengebracht in een globale blauwdruk, een piramidemodel, waarin toetsinhoud, domeinen en aandoeeningen, en toetsniveaus, van kennen en kunnen tot (dagelijks) doen, onderscheiden worden. Dit model kan behulpzaam zijn bij het ontwikkelen van toetsen voor verschillende onderwerpen en fases van de geneeskundeopleidingen. De voorspellende waarde van kennis- en vaardigheidstoetsen voor de kwaliteit van feitelijk handelen in de praktijk blijkt bij praktiserende huisartsen gering. Kennis en kunde kunnen in de eerste opleidingsfasen door kennistoetsen en vaardigheidstoetsen worden beoordeeld. Uiteindelijk heeft toetsing door observatie van het handelen van studenten en artsen in de klinische praktijk de voorkeur: daar moeten immers kennis en kunde integraal worden toegepast.

Directe observatie: Directe observatie in de klinische praktijk vereist standaardisatie van de aangeboden casuïstiek, gezien de grote variatie in patiënten en aandoeeningen. Standaardisatie bevordert de validiteit, betrouwbaarheid en haalbaarheid van de methode. Selectie van casus is een belangrijk instrument. Deze wordt beschreven, evenals de noodzakelijke betrouwbaarheidsanalyse en kostenaspecten. Twee methoden van directe observatie van het handelen in de klinische praktijk worden kort uitgewerkt: observatie door incognito (simulatie)patiënten en video-observatie. (Ram P, Hobma S, Aretz K. Viva in vivo assessment (VIVA); toetsing van het handelen van artsen en studenten door observatie in de klinische praktijk. Tijdschrift voor Medisch Onderwijs 2001;20(5):196-204.)

Inleiding

Een belangrijk aspect van het medisch (post)academisch onderwijs is het implementeren van toetsen waarmee de kwaliteit van handelen van artsen en studenten gemeten kan worden.^{1 2} Het brede medische vakgebied kan niet door een enkele toets gedekt worden. Er zijn verschillende toetsen beschikbaar die, eventueel in onderlinge samenhang, kunnen worden toegepast.³⁻⁵ Deze rijkdom aan toetsmogelijkheden roept de vraag op wat op welk niveau met welke methoden getoetst moet worden. Communicatieve aspecten bijvoorbeeld kunnen getoetst worden op

het niveau van kennis of vaardigheden, maar ook op het niveau van feitelijk handelen in de klinische praktijk.

Directe observatie van het handelen van studenten en artsen in de klinische praktijk is aantrekkelijk, want in de praktijk moeten kennis en kunde uiteindelijk in contacten met echte patiënten en hun context worden toegepast. Uit oogpunt van contextueel leren pleit men dan ook voor het zo snel mogelijk inpassen in de opleiding van leermomenten met echte patiënten.⁶ Het is logisch om dit 'praktijkleren' dan ook in de praktijk te toetsen. Een inhoudelijke congruentie van onder-

wijs en toetsing motiveert studenten bovendien om meer praktijkgericht, dus patiënt- en contextgericht, te leren. Toetsing stuurt immers het leerproces.² Directe observatie van het handelen van artsen in de praktijk van alledag geeft inzicht in wat goed gaat in de patiëntenzorg en wat verbetering behoeft. Men kan daarmee de inhoud en de effectiviteit van nascholing vaststellen. Indirecte methoden, zoals zelfregistratie, het analyseren van prescriptiecijfers of gegevens betreffende aanvullende diagnostiek, zijn goede alternatieven, maar minder toepasbaar bij studenten.⁷

Directe observatie in de klinische praktijk heeft een aantal praktische beperkingen. De praktijk van alledag is grillig en niet gestandaardiseerd, hetgeen een betrouwbare vergelijking van studenten of artsen bemoeilijkt.⁸ Er moeten veel consulten getoetst worden om tot een algemeen oordeel te kunnen komen. Immers, de score behaald voor de kwaliteit van handelen in het ene consult voorspelt de kwaliteit van handelen in een andersoortig consult slechts matig.⁹ Deze 'casusspecificiteit' maakt deze toetsing bewerkelijk en kostbaar. Een ander probleem betreft de selectie van te toetsen onderwerpen. Dagelijkse consulten zijn allemaal anders. De casuïstiek in de praktijk is eerder complex dan eenvoudig. Welke consulten komen voor toetsing in aanmerking en welke niet? Zijn de geselecteerde consulten representatief voor de te toetsen onderwerpen en taken, en zijn de steekproeven vergelijkbaar? Bij toetsing in de praktijk is de privacy van de patiënt in het geding.¹⁰ Toestemming van de patiënt voor observatie is een vereiste. Dit kan de selectie van consulten, en dus de inhoudsvaliditeit van de toets, beïnvloeden.

Mede door deze beperkingen van 'in vivo' toetsing, worden kennistoetsen en 'in vitro' simulatietoetsen, in (post)academisch

onderwijs veel gebruikt. Kennistoetsen zijn minder bewerkelijk, op grote schaal toepasbaar en gestandaardiseerd. De validiteit neemt toe als de kennistoets patiëntgeoriënteerd is en realistische praktijksituaties weerspiegelt.¹¹ De voorspellende waarde van kennistoetsscores voor de kwaliteit van het dagelijks handelen van artsen blijkt echter beperkt.^{12 13} Het succesvol scoren op een kennistoets is nog geen garantie voor het adequaat stellen van een diagnose, het uitwerken van een goed beleid, en het inventariseren en beantwoorden van hulpvragen in de dagelijkse praktijk.

Een vaardigheidstoets waarin men een bepaalde taak moet uitvoeren, benadert de werkelijkheid beter. Deze methode is gericht op observeerbaar gedrag en is gestandaardiseerd. Iedereen krijgt dezelfde opdracht in een identieke setting, bijvoorbeeld in een skillslab met (simulatie)patiënten. Deze 'in vitro' methode is wel bewerkelijk. De voorspellende waarde van simulatietoetsen voor het handelen van artsen in de dagelijkse praktijk is beperkt. 'Laten zien wat je kunt' en 'dagelijks doen' zijn mogelijk verschillende constructen.^{13 14}

Competentieontwikkeling is de essentie van medisch onderwijs. De kwaliteit ervan kan men beoordelen met kennis- en vaardigheidstoetsen. Naast deze competentietoetsen zijn ook valide en betrouwbare 'in vivo' toetsen nodig om de kwaliteit van het handelen van de student in de klinische realiteit te beoordelen. Door toetsing in de kliniek kan men bovendien de samenhang tussen kennis en vaardigheden en de kwaliteit van de uiteindelijke toepassing ervan in de praktijk vaststellen. Toetsen in de klinische realiteit vraagt om specifieke activiteiten om de validiteit en betrouwbaarheid van de toets te bevorderen.

In het navolgende wordt een model beschreven dat inzicht biedt in de overwe-

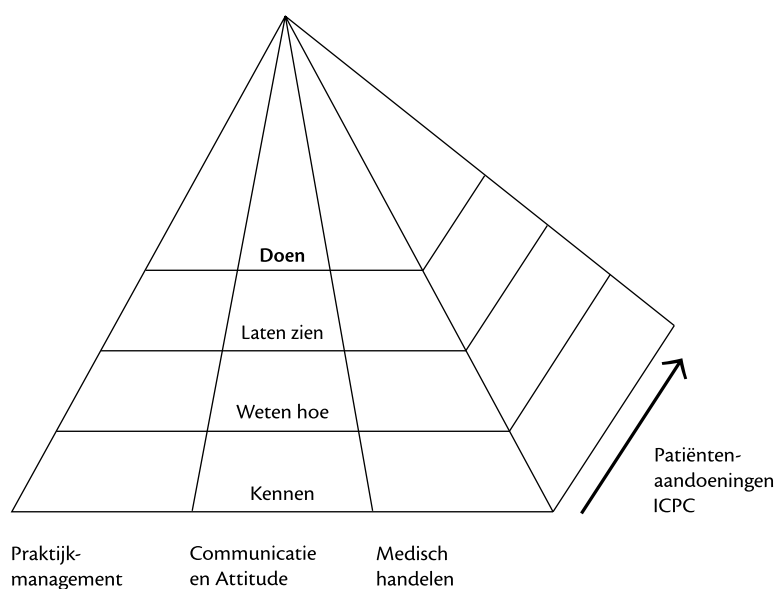
gingen bij het kiezen van een toetsvorm. Vervolgens worden belangrijke psychometrische aspecten van 'in vivo' toetsing behandeld. Standaardisatie is een belangrijke vereiste bij deze toetsen. Twee soorten gestandaardiseerde observatietoetsen worden besproken: toetsing met behulp van 1) gestandaardiseerde patiënten en 2) video-opnames.

Het gebruik van een globale blauwdruk: een piramidemodel

Figuur 1 toont een globale blauwdruk voor toetsing. Deze is ontwikkeld uit drie verschillende modellen: het model van Fabb en Marshall, het model van Miller en dat van Donabedian.¹⁵⁻¹⁷ Het model van Fabb en Marshall beschrijft de domeinen van de (huisarts)geneeskunde, inclusief aandoeningen en categorieën patiënten. Het model van Miller weerspiegelt de hiërarchische ontwikkeling van medische competentie, via kennen, weten

hoe kennis toe te passen (klinisch redeneren), kunnen (vaardig zijn) tot feitelijk doen in de klinische praktijk. Het model van Donabedian is een formule: arts-patiëntcommunicatie en vakinhoudelijk handelen (proces) en praktijkmanagement (structuur) leiden tot patiëntuitkomsten (gezondheid en tevredenheid). Deze drie modellen zijn samengevoegd tot een piramidemodel, een globale blauwdruk voor toetsing.

De onderlinge samenhang van toetscores voor praktijkmanagement, arts-patiëntcommunicatie en medisch handelen in de praktijk ('doen') is voor huisartsen onderzocht. De kwaliteit van praktijkmanagement correleert vrijwel niet met het communicatief en medisch handelen van de huisarts in de spreekkamer. Het beoordelen van verschillende aspecten van praktijkmanagement, zoals registratie of instrumentarium (structuur), geeft dus weinig informatie over het handelen van



Figuur 1. Een globale blauwdruk voor toetsing.

de huisarts in patiëntcontacten (proces).¹⁸ Dat dient derhalve zo direct mogelijk getoetst te worden, bijvoorbeeld door observatie van consulten. Het toetsen van kennis en vaardigheden van (huis)artsen 'voorspelt' het feitelijk handelen in de directe patiëntenzorg redelijk, maar onvoldoende om toetsing daarvan te kunnen vervangen. Dit is aangetoond voor arts-patiëntcommunicatie en medisch handelen.^{5 13 19} Bij directe observatie van echte consulten blijkt dat arts-patiëntcommunicatie en medisch handelen vrij sterk samenhangen.²⁰

Deze onderzoeksresultaten kan men niet generaliseren naar het basiscurriculum. Onderzoek zal moeten aantonen in welke mate kennis- en vaardigheidstoetsen het gedrag van studenten in de kliniek kunnen voorspellen. Valide en betrouwbare toetsing van het handelen van de studenten in de kliniek is daarvoor een voorwaarde.

Het domein praktijkmanagement is in het basiscurriculum beperkt, maar belangrijke aspecten ervan zoals hygiëne, onderhoud en gebruik van instrumentarium, registratie van gegevens of samenwerking met derden kunnen in dit domein worden opgenomen. Attitude is binnen alle domeinen aan de orde, maar komt vooral in arts-patiëntcommunicatie tot uiting.²¹ De hiërarchie van kennen, weten hoe, kunnen en doen is bij uitstek van toepassing op het basiscurriculum. Men kan per opleidingsfase aangeven in welke mate en voor welke aspecten kennistoetsen, vaardigheidstoetsen of 'in vivo' toetsen geïndiceerd zijn.

In het piramidemodel zijn op de horizontale as praktijkmanagement, arts-patiëntcommunicatie, attitude en medisch handelen opgenomen. Op de diepte-as zijn de categorieën patiënten (zoals zuigelingen of ouderen) en aandoeningen weergegeven, met het internationale classificatiesysteem voor de huisarts-

geneeskunde (ICPC) als voorbeeld. Op deze assen kunnen onderwijsdoelen, eindtermen en items uit het raamcurriculum worden opgenomen. De verticale as weerspiegelt vier te onderscheiden niveaus: van kennen, via weten hoe en kunnen, naar doen, het feitelijk handelen.

Wat zijn de voordelen van dit model? Bij het selecteren van te toetsen onderwerpen wordt men gedwongen te formuleren op welk niveau men deze wil toetsen en waarom. De toetsbaarheid van onderwerpen op het niveau van feitelijk handelen kan beperkt zijn door de lage incidentie van de aandoening (leukemie), de hoge mate van spoed (reanimatie) of de mate waarin de patiënt belast kan en mag worden (terminale zorg). Kennistoetsen of vaardigheidstoetsen zijn in dat geval betere alternatieven. Men kan bovendien nagaan of evenwichtig aandacht is besteed aan toetsing van verschillende domeinen in relatie tot aandoeningen, patiëntencategorieën en te onderscheiden competentieniveaus, gegeven de fase van de opleiding. Bij combinatie van verschillende toetsmethoden maakt dit model inzichtelijk of verschillende onderwerpen binnen een domein op verschillende niveaus worden getoetst en of verschillende onderwerpen binnen verschillende domeinen aan de orde komen. Het model dwingt tot keuzes en ondersteunt de eerste globale beschrijving van de toetsinhoud en daarmee de inhoudsvaliditeit.

Belangrijke psychometrische aspecten van 'in vivo' toetsing

De inhoudsvaliditeit van 'in vivo' toetsing (meet men wat men beoogt te meten) wordt vooral bepaald door het score-instrument en de inhoud van de geselecteerde consulten. In de klinische praktijk staat de integratie van technische en communicatieve vaardigheden centraal. Een scorelijst met globale items heeft daarom

de voorkeur boven een checklist met gedetailleerde items.²² Als men de werkelijkheid van de klinische praktijk fragmenteert in op zichzelf staande gedetailleerde competenties, wordt de integratie ervan niet gemeten. De inhoudsvaliditeit van de toetsmethode wordt ook bepaald door de inhoud van de consulten. In de consulten moeten die klinische taken aan de orde zijn die men beoogt te meten. In de dagelijkse praktijk weet men echter vaak niet van tevoren waar de patiënt voor komt. Er zijn een aantal oplossingen. Gestandaardiseerde (simulatie)patiënten kunnen 'incognito' het spreekuur bezoeken en tijdens het consult de arts of student observeren.²³ Men kan ook selectief video-opnames van consulten maken of na opnames consulten selecteren. Criteria waaraan de casus moeten voldoen zijn nodig.^{20 24}

Constructvaliditeit is de mate waarin een toetsmethode in staat is een hypothese (construct) aan te tonen, bijvoorbeeld een verschil tussen scores van beginners en ervaren artsen. De lengte en de moeilijkheidsgraad van consulten kunnen de scores beïnvloeden en daarmee de constructvaliditeit.²⁰ Wil men twee steekproeven met elkaar vergelijken, dan dienen deze vergelijkbaar te zijn. Ook daarom zijn criteria waaraan de casus moeten voldoen nodig.

Het beperken van ongewenste variantiebronnen bevordert de consistentie, dus de betrouwbaarheid, van de meting. Er moeten heldere procedures zijn voor het trainen van observatoren en (simulatie)patiënten, de installatie van de video-apparatuur, het vragen van toestemming aan de patiënt en de selectie van consulten. Kortom, standaardisatie is een vereiste. De betrouwbaarheid dient bij voorkeur vastgesteld te worden met behulp van generaliseerbaarheidsanalyse, gezien het grote aantal variantiebronnen.²⁵

De haalbaarheid van de methode, de mate van belasting voor organisatoren en deelnemers, wordt bepaald door de financiële mogelijkheden en de educatieve impact ('de winst') voor de deelnemers. De kosten van simulatietoetsen zijn hoog.²⁶ Toetsing van echte consulten in de klinische praktijk heeft een aantal financiële voordelen. Er zijn geen kosten voor het schrijven en valideren van simulatiecasus, voor reiskosten en training van simulatiepatiënten, huur van ruimtes in een skillslab en de inrichting ervan (instrumentarium en verbruiksgoederen).

De educatieve impact van toetsing in de praktijk doet niet onder voor die van competentietoetsen. Afwerende argumenten zoals 'een simulatie is gekunsteld, de praktijk is heel anders', 'in het echt doe ik het veel beter' en 'die kennistoets meet theorie' zijn niet van toepassing. Huisartsen waarderen feedback op basis van praktijktoetsen meer dan feedback op basis van simulatie- en kennistoetsen. Belangrijk argument daarbij is de natuurgetroouwheid van de toetsinhoud.⁵

Een ordening van de complexe werkelijkheid van alledag is nodig om een goede beoordeling mogelijk te maken. Er zijn verschillende mogelijkheden om toetsing van het handelen van artsen en studenten in de klinische praktijk te standaardiseren.

Observatie door incognito (simulatie)patiënten

Het vastleggen van het feitelijk handelen van studenten of artsen door het introduceren van incognito gestandaardiseerde patiënten (SP's) in spreekuren in de polikliniek of huisartsenpraktijk blijkt een valide en betrouwbare observatiemethode.²⁷ ²⁸ SP's zijn echte patiënten of gezonde personen die getraind zijn om een aandoening accuraat en consistent te spelen. Zij bezoeken de arts of student in de klinische praktijk en observeren en registreren het

handelen van de arts of student. De gesimuleerde persoonsgegevens en medische voorgeschiedenis van de SP zijn levensecht in de praktijk geregistreerd. De SP is daarin niet te onderscheiden van een echte patiënt. De arts of student heeft vooraf toestemming gegeven voor deze procedure, maar weet niet door wie en wanneer hij geconsulteerd wordt. In principe kunnen alle onbekende patiënten SP's zijn. Tijdens de training van SP's wordt veel aandacht geschonken aan het zorgvuldig spelen en betrouwbaar weergeven van de rol door het invullen van casusspecifieke checklists door de SP en de trainer. Pas na voldoende overeenkomst tussen de scores van de SP en de trainer, begint de SP met de 'undercover' praktijkbezoeken. Het percentage dat 'ontmaskerd' wordt bedraagt 10-20%. Het voordeel van deze methode is de grote mate van standaardisatie: iedere deelnemer ziet dezelfde patiënt, die consistent hetzelfde probleem voorlegt en het handelen consistent scoort.

Observatie van videoregistraties van echte consulten

Vergeleken met directe observaties door een in de spreekkamer aanwezige observator zijn videoregistraties minder storend. Video-opnames kunnen door meerdere observatoren worden beoordeeld. Zij kunnen individueel of in een onderwijsgroep worden bekeken ter ondersteuning van de feedback; en observeren is leerzaam. Medisch technische vaardigheden komen echter vaak niet in detail in beeld. De procedure is als volgt.

1. Een video-observatiesysteem wordt geïnstalleerd in de praktijk, bij voorkeur met twee camera's en microfoons. Daarmee worden consulten volledig opgenomen, inclusief het lichamelijk onderzoek. In kleinere ruimtes voldoet een draaibare camera. Met slechts een vaste camera in de spreekkamer mist men het lichamelijk onderzoek. En het gaat juist om observatie van integratie van kennis en vaardigheden.
2. Het vragen van toestemming aan de patiënt (informed consent) kan het best schriftelijk afgehandeld worden aan de balie: de patiënt heeft tijd om na te denken en het consult begint ongestoord.²⁹ Hoe natuurlijker de arts of student werkt, des te minder invloed heeft de observatie op diens gedrag. Dit beperkt het 'audience effect'.^{30 31} De patiënt krijgt van de assistente een groen of rood kaartje (wel of geen toestemming) en geeft dat ondertekend aan de arts of student. Bij toestemming start deze de opname. Tijdens het consult en erna kan de patiënt de toestemming intrekken door dit kenbaar te maken aan de arts. Deze wist dan de opname direct. De toestemming kan ook op video worden opgenomen aan het begin van het consult. In ieder geval moet de procedure zwart op wit staan.
3. Na het consult noteert de arts of student consultgegevens in een logboek, zoals (aantal) klachten, werkhypothese, geslacht, leeftijd, duur en aard van het consult (initieel of vervolg). Uit dit logboek selecteert de toetscommissie of observator een aantal consulten aan de hand van bepaalde criteria.
4. Deze criteria zijn tevoren opgesteld en zijn afhankelijk van het doel van de toetsing. Wil men een algemene uitspraak doen over de kwaliteit van het handelen van de arts of student, dan moet een representatieve steekproef van consulten samengesteld worden van voldoende omvang om tot generaliseerbare scores te komen. Voor de huisartsgeneeskunde zijn hiervoor selectiecriteria vastgesteld (tabel 1). Deze criteria zijn afgeleid van preva-

Tabel 1. *Criteria voor de selectie van consulten (16 per huisarts) en per criterium het percentage steekproeven dat hieraan voldoet (videotoetsing huisartsen).*

Criteria	% steekproeven dat voldoet aan criterium
1. Minimaal 8 consulten uit verschillende ICPC-hoofdstukken ¹	100.0
2. De eerste 5 opgenomen consulten zijn niet geïncludeerd	94.6
3. 8 consulten zijn te relateren aan NHG-standaarden ²	86.2
4. Representativiteit van de ICPC-hoofdstukken: D, K, L, R en S	71.3
5. Minimaal 14 consulten met een duur van >5 en <15 minuten/consult	86.2
6. Leeftijdverdeling: jongste patiënt <18 jaar en de oudste >65 jaar	84.0
7. Man-vrouw patiënten: ≥ 6 en ≤10 vrouwelijke patiënten	72.3
8. Ten minste 8 initiële consulten, ten minste 4 vervolconsulten	70.2
9. Maximaal 2 consulten met >2 redenen van komst	92.6
10. Maximaal 2 consulten met uitsluitend psychosociale problemen	93.6

¹ ICPC: International Classification of Primary Care

² NHG: Nederlands Huisartsen Genootschap

lenticijfers, taken, standaarden en richtlijnen voor huisartsgeneeskundig handelen en toetsgegevens, zoals 'audience effect' en 'casusspecificiteit'.²⁰

- De te observeren consulten worden verdeeld over de observatoren. Een observator scoort een beperkt aantal consulten om te voorkomen dat de ene observatie doorwerkt in de daarop volgende (halo-effect). De gemiddelde scores per consult, per gemeten vaardigheid (taak) en per steekproef worden berekend (data-invoer en data-analyse).
- De steekproefkenmerken per deelnemer worden geanalyseerd. In hoeverre zijn er verschillen tussen deelnemers? Wordt de ene deelnemer aanmerkelijk korter geobserveerd dan de ander? Is er voldoende variatie aan onderwerpen? Kortom, in welke mate zijn de selectiecriteria geëffectueerd? Met een variantieanalyse wordt de invloed van de steekproefverschillen op de scores verklaard, bijvoorbeeld door een multiple regressieanalyse met de scores als afhankelijke variabelen en de

steekproefkenmerken als onafhankelijke variabelen. Deze analyse is obli-gaat, omdat er nog onvoldoende bekend is over de invloed van deze variantiebronnen. Niet alle steekproef-verschillen beïnvloeden de validiteit in negatieve zin. Binnen bepaalde grenzen is bijvoorbeeld de tijd die de arts of student aan de patiënt besteedt een kenmerk van kwaliteit.²⁰

Conclusie

Onderwijskundige en psychometrische argumenten pleiten voor observatie van het handelen van studenten en artsen in de klinische praktijk. Observatie in de praktijk is niet bewerkelijker dan observatie van vaardigheden in een simulatieset-ting. Het gebruik van globale score-instru-menten, een zorgvuldige standaardisatie, een analyse van de steekproefkenmerken (bij selectie van echte consulten) en een adequate betrouwbaarheidsanalyse zijn essentiële voorwaarden voor een goede acceptatie en educatieve impact.

Voor het basiscurriculum kunnen onder-zoeksresultaten bij praktiserende (huis)-

artsen een uitdaging zijn om valide en betrouwbare toetsing in de kliniek te bevorderen. Het mes snijdt dan aan twee kanten: de 'voorspellende waarde' van kennis- en vaardigheidstoetsen voor de kwaliteit van handelen van studenten in de echte patiëntenzorg wordt inzichtelijker en het medisch onderwijs wordt belangrijke educatieve en selectieve methoden rijker.

Literatuur

- Grol R, Wensing M, Jacobs A, Baker R. Quality assurance in general practice. The state of the art in Europe. Utrecht: NHG and EquiP; 1993.
- Vleuten CPM van der. The assessment of professional competence: developments, research and practical implications. *Advances in Health Sciences Education* 1996;1:41-67.
- Verwijnen GM, Pollemans MC, Wijnen WHFW. Voortgangstoetsing. In: Metz JCM, Scherpbier AJJA, Vleuten CPM van der, redactie. *Medisch onderwijs in de praktijk*. Assen: Van Gorcum; 1995. p. 225-231.
- Pollemans M. Kennistoetsing bij huisartsen [dissertatie]. Maastricht: Universitaire Pers Maastricht; 1994.
- Ram P. Comprehensive assessment of general practitioners: a study on validity, reliability and feasibility [dissertatie]. Maastricht: Unigraphic; 1998.
- Rethans JJ, Westin S, Hays R. Methods for quality assessment in general practice. *Fam Pract* 1996;5:468-76.
- Scherpbier AJJA, Verwijnen GM, Schaper N, Dunselman GAJ, Vleuten CPM van der. Vaardigheidsonderwijs nu en in de toekomst. *Tijdschrift voor Medisch Onderwijs* 2000;19(1):6-16.
- Kane MT. The assessment of professional competence. *Evaluation & the Health Professions* 1992;15:163-82.
- Elstein A, Shulman LS, Sprafka SA. *Medical problemsolving: An analysis of clinical reasoning*. Cambridge Massachusetts: Harvard University Press; 1978.
- Shafir MS. Patient consent to observation. *Can Fam Phy* 1995;41:1367-72.
- Leeuwen YD van, Pollemans MC, Mol SS, et al. The Dutch knowledge test for general practice: issues of validity. *Eur Jour Gen Pract* 1995; 1:113-7.
- Ramsey PG, Carline JD, Inui YS et al. Predictive validity of certification by the American Board of Internal Medicine. *Ann Int Med* 1989;110:719-26.
- Ram P, Vleuten CPM van der, Rethans J-J, Schouten B, Hobma S, Grol R. Assessment in General Practice: the predictive value of written knowledge tests and a multiple station examination for actual medical performance in daily practice. *Med Educ* 1999;33:197-203.
- Rethans JJ. Does competence predict performance? standardized patients as a means to investigate the relationship between competence and performance of general practitioners [dissertatie]. Amsterdam: Thesis Publ; 1991.
- Fabb WE, Marshall JR. *The assessment of clinical competence in general family practice*. Lancaster, England and Hingham, USA: MTP Press Limited; 1983.
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;65:S63-S67.
- Donabedian A. Exploration in quality assessment and monitoring 1. The definition of quality and approaches to its assessment. *Ann Arbor: Health Administration Press*; 1980.
- Ram P, Hombergh P van den, Grol R, Rethans J-J, Vleuten C van der. Structuur en proces. Een onderzoek naar de samenhang tussen praktijkvoering en medisch en communicatief handelen door huisartsen. *Huisarts Wet* 1999;42:493-8.
- Ram P, Rethans J-J, Grol R, Vleuten C van der, Hobma S, Aretz K. Integrale toetsing: de voorspellende waarde van een simulatiespreekuur en kennistoetsen voor de kwaliteit van het dagelijks handelen. *Huisarts Wet* 2000;43:103-10.
- Ram P, Grol R, Rethans J-J, Vleuten C van der. Videotoetsing van consulten van huisartsen in de eigen praktijk. een onderzoek naar validiteit, betrouwbaarheid en haalbaarheid. *Huisarts Wet* 1999;42:439-45.
- Luijk SJ van, Santen-Hoeufft M van, Hillen HFP, Smeets JGE. Het beoordelen van professioneel gedrag bij studenten en artsen. *Tijdschrift voor Medisch Onderwijs* 2000;19(5):163-72.
- Luijk SJ van, Vleuten CPM van der. Het stations-examen. In: Metz JCM, Scherpbier AJJA, Vleuten CPM van der, redactie. *Medisch onderwijs in de praktijk*. Assen: Van Gorcum & Comp.; 1995. p. 202-207.
- Rethans JJ, Sturmans F, Drop MJ, Vleuten CPM van der. Assessment of performance in actual practice of general practitioners by use of standardized patients. *Br J Gen Pract* 1991;41:97-9.
- Campbell LM, Howie JGR, Murray TS. Use of videotaped consultations in summative assessment of trainees in general practice. *Br J Gen Pract* 1995;45:137-41.
- Vleuten CPM van der, Wijnen WHFW. Niets praktischer dan een goede theorie: generaliseerbaarheidstheorie als instrument voor betrouwbaarheidsstudies. *Bulletin Medisch Onderwijs* 1991; 10:2-14.
- Reznick RK, Smee S, Baumber JS et al. Guidelines for estimating real cost of an objective structured clinical examination. *Acad Med* 1993;68:513-7.

27. Rethans J-J. Needs assessment in continuing medical education through standardized patients. *J Cont Educ Health Prof* 1998;18:172-8.
28. Collins JP, Harden RM. Real patients, simulated patients and simulators in clinical examinations. *Med Teach* 1998;20:508-21.
29. Shafir MS. Patient consent to observation. *Can Fam Phy* 1995;41:1367-72.
30. Wakefield J. Direct observation. In: Neufeld VR, Norman GR, editors. *Assessing clinical competence*. New York: Springer Publishing Company; 1985. p. 51-71.
31. Pringle M, Stewart-Evans C. Does awareness of being video recorded affect doctor's consultation behaviour? *Br J Gen Pract* 1990;40:45-8.

De auteurs:

Paul Ram, Sjoerd Hobma en Karin Aretz, Werkgroep Onderzoek Kwaliteit (WOK) van de Katholieke Universiteit Nijmegen, de Universiteit Maastricht en het Nederlands Huisartsen Genootschap (NHG) te Utrecht.

Correspondentieadres:

Paul Ram, huisarts-stafid, Universiteit Maastricht/ Nederlands Huisartsen Genootschap, Capaciteitsgroep Huisartsgeneeskunde, Universiteit Maastricht, Postbus 616, 6200 MD Maastricht, Paul.Ram@HAG.unimaas.nl.

Summary

Introduction: Adequate assessment of clinical competence requires the use of different types of test methods. The choice of test depends on what is to be assessed and at which level of training.

Blueprint: Key aspects to be considered in any choice of test format in medical undergraduate and postgraduate training have been combined in a global blueprint, shaped like a pyramid and comprising test content, domains and disorders, and level of assessment – from knows, knows how, and shows how to does. The model can help to select the most appropriate test for a specific subject at a specific level of training. Knowledge and skills tests have been found to have only limited predictive value for GPs' actual performance in clinical practice. In the early stages of medical training competence can be assessed using knowledge and skills tests. Eventually, the preferred assessment method is direct observation of students and physicians applying their knowledge and skills in an integrated manner in actual clinical practice.

Direct observation: Testing through direct observation in clinical practice requires standardized cases, given the tremendous variety in patients and disorders in everyday clinical practice. Standardization promotes validity, reliability and feasibility of the assessment. Selection of cases is described as well as analysis of reliability and costs. Two observational test formats are described in some detail: observation by incognito standardized patients and video-observation. (Ram P, Hobma S, Aretz K. Viva in vivo assessment (VIVA); assessment of the performance of students and doctors in clinical practice. *Dutch Journal of Medical Education* 2001;20(5):196-204.)